Scaling First Principles Materials and Nanoscience Codes to Thousands of Processors (and Thousands of Atoms)

(Plane-wave, DFT codes)

Andrew Canning and Lin-Wang Wang

Computational Research Division

LBNL

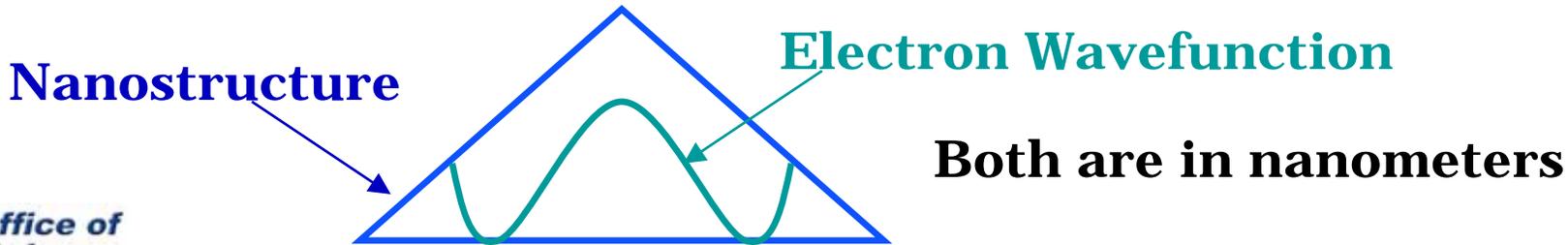**Definition:** **Nanostructure is an assembly of nanometer scale "building blocks".**

Why nanometer scale: This is the scale when the properties of these "building blocks" become different from bulk.

size

Nanostructure

Electron Wavefunction
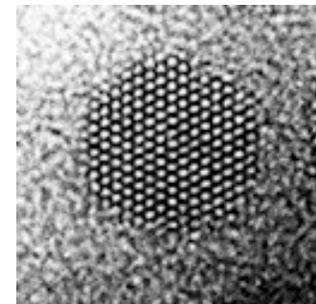
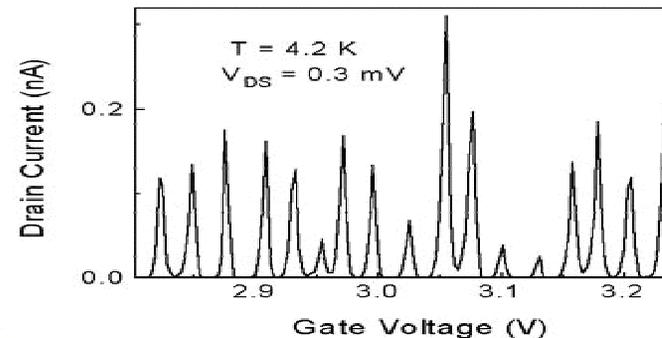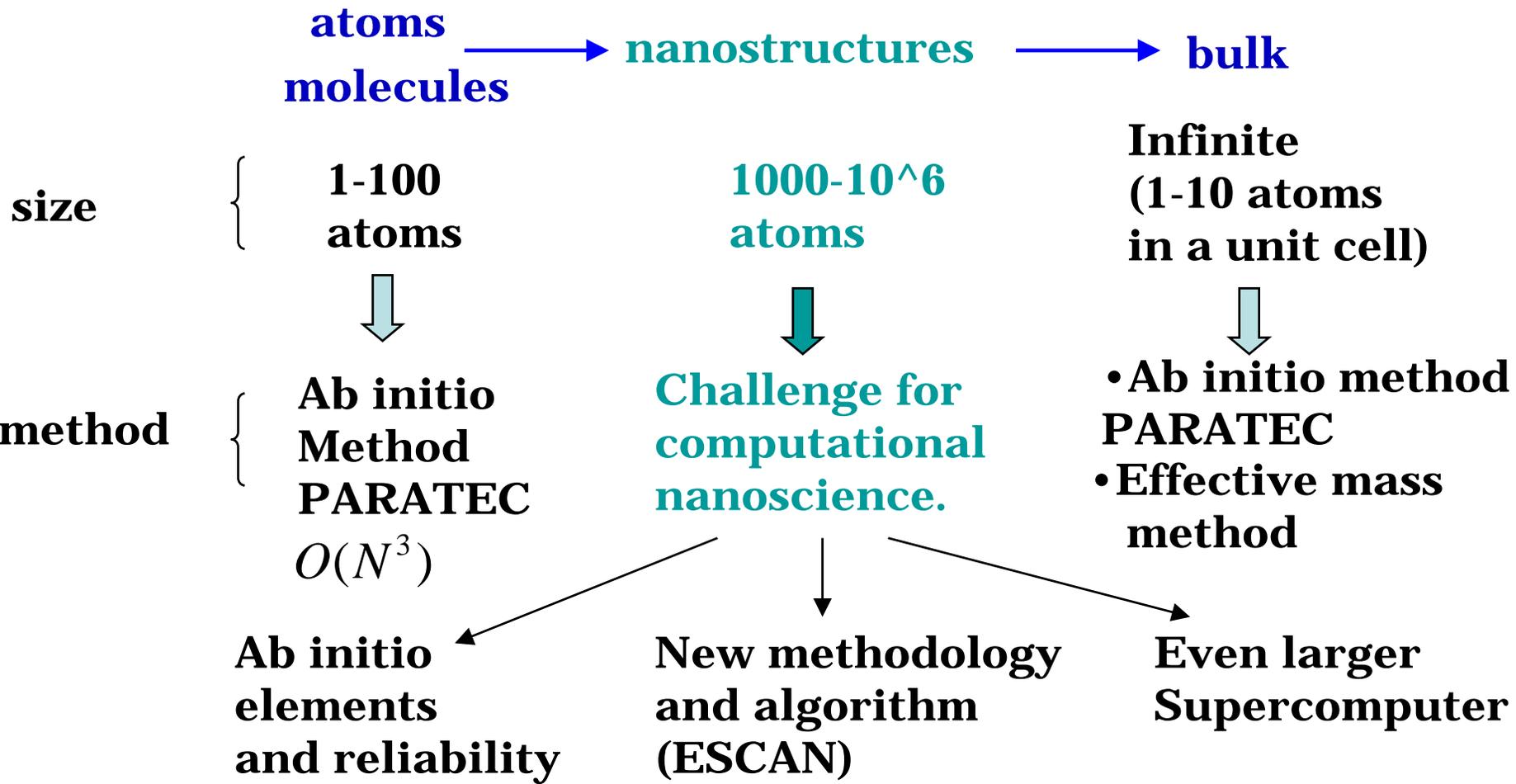**Both are in nanometers**

CRD

- **Band gap increase**

- **Single electron effects on transport (Coulomb blockade).**

- **Mechanical properties, surface effects and no dislocations**

CdSe Quantum Dots (Bawendi Group)

Photo by Felice Frankel

CdSe quantum dot

T = 4.2 K
V_{DS} = 0.3 mV

Drain Current (nA)
Gate Voltage (V)

atoms molecules → nanostructures → bulk

size			
	1-100 atoms	1000-10^6 atoms	Infinite (1-10 atoms in a unit cell)

method			
	Ab initio Method PARATEC $O(N^3)$	Challenge for computational nanoscience.	• Ab initio method PARATEC • Effective mass method

Ab initio elements and reliability

New methodology and algorithm (ESCAN)

Even larger Supercomputer

Office of Science
U.S. DEPARTMENT OF ENERGY

$$\{-\frac{1}{2}\nabla^2 + \int \frac{\rho(r')}{|r-r'|}dr' + \sum_I \frac{Z}{|r-R_I|} + V_{XC}(r)\}\psi_j(r) = E_j\psi_j(r)$$

Solve Kohn-Sham Equations self-consistently for electron wavefunctions

1. **Plane-wave expansion for** $\psi_{j,k}(r) = \sum_g C_g^j(k)e^{i(g+k).r}$

2. **Replace "frozen" core by a pseudopotential**

   Different parts of the Hamiltonian calculated in different spaces (fourier and real) 3d FFT used

- **Divide sphere of plane-waves in columns between processors (planes of grid in real space)**

- **Parallel 3d FFT used to move between real space and fourier space**

- **FFT requires global communications data packet size ~ 1/(# processors)^2**

(a) (b) (c) (d) (e) (f)
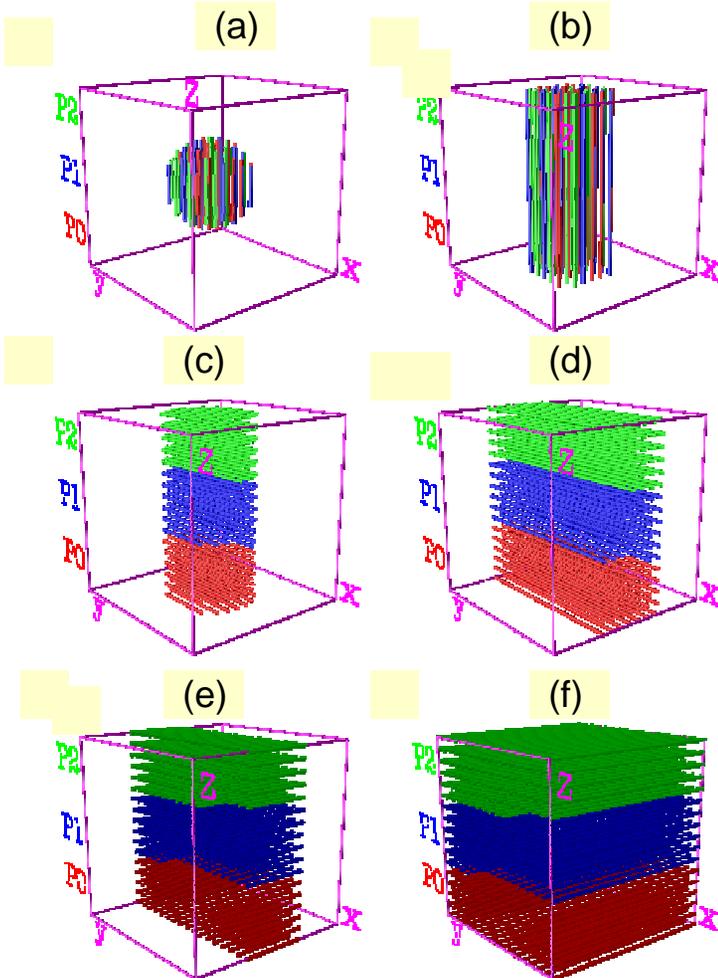
- 3D FFT done via 3 sets of 1D FFTs and 2 transposes

- Most communication in global transpose (b) to (c) little communication (d) to (e)

- Flops/Comms ~ logN

- Many FFTs done at the same time to avoid latency issues

- Only non-zero elements communicated/calculated

- Much faster than vendor supplied 3D-FFT

- **Minimize the global communications part**

- **Latency problem: Use all-band method in conjunction with many FFTs at the same time to make data packets larger (new all band CG method for metals)**

- **FFT part scales as $N^2 logN$ while other parts scale as N^3 (larger systems scale better)**

Parallel Fourier techniques (many-band approach) used in the following codes:

- **PARATEC PARAllel Total Energy Code (plane-wave pseudopotential code)**
- **PEtot (plane-wave ultrasoft pseudopotential code)**
- **ESCAN (Energy SCAN) Uses folded spectrum method for non-selfconsistent nanoscale calculations with plane-waves for larger systems.**
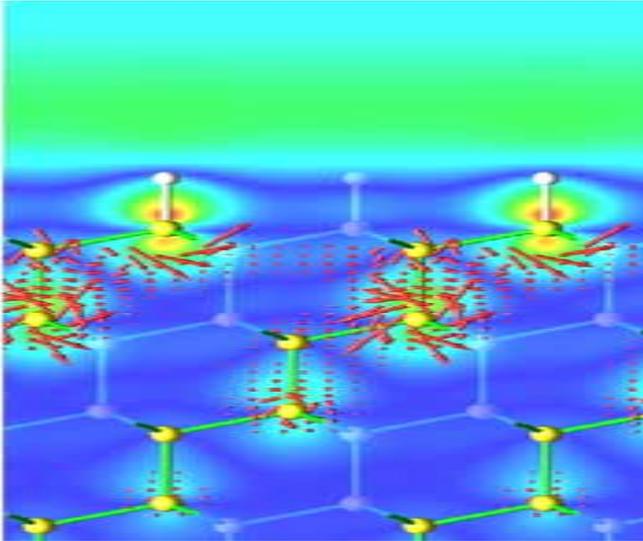
- **PARATEC performs first-principles quantum mechanical total energy calculation using pseudopotentials & plane wave basis set**
- **Designed to run on large parallel machines   IBM SP etc. but also runs on PCs**

- **PARATEC uses all-band CG approach to obtain wavefunctions of electrons**
- **Generally obtains high percentage of peak on different platforms**
- **Developed with Louie and Cohen's groups (UCB, LBNL), Raczkowski (Multiple 3d FFTs   Peter Haynes and Michel Cote)**

PARATEC: Performance

P	IBM SP Power 3		IBM SP Power4		SGI Altix		NEC ES		CRAY X1	
	Gflops/P	%peak	Gflops/P	%peak	Gflops/P	%peak	Gflops/P	%peak	Gflops/P	%peak
32	0.950	63%	2.02	39%	3.71	62%	4.76	60%	3.04	24%
64	0.848	57%	1.73	33%	3.24	54%	4.67	59%	2.59	20%
128	0.739	49%	1.50	29%			4.74	59%	1.91	15%
256	0.572	38%	1.08	21%			4.17	52%		
512	0.413	28%					3.39	42%		
1024							2.08	26%		

432 Atom Si Bulk System   5 CG step

- **First time code achieved over 1 Tflop
  Aggregate 2.6 TFlops for 686 Si atom**

- **Previous best was .7 TFlop on Power3 using 1500 procs**

Work carried out with L. Oliker (CRD) J.T. Carter (NERSC)

PARATEC: Performance on ES SX6

P	NEC SX6 ES			CRAY X1		
	Gflops/P	Gflops	%peak	Gflops/P	Gflops	%peak
64	5.25	335	66%	3.73	239	29%
128	4.95	633	62%	3.01	385	24%
256	4.59	1176	57%			
512	3.76	1924	47%			
1024	2.53	2591	32%			

Scaling Issues 686 atom Si

- **% time in communications increasing (3D FFT)**
- **Fewer number of multiple 1D FFTs for each proc. (vector length drops)**
- **Matrices for BLAS3 becoming smaller (vector length drops)**
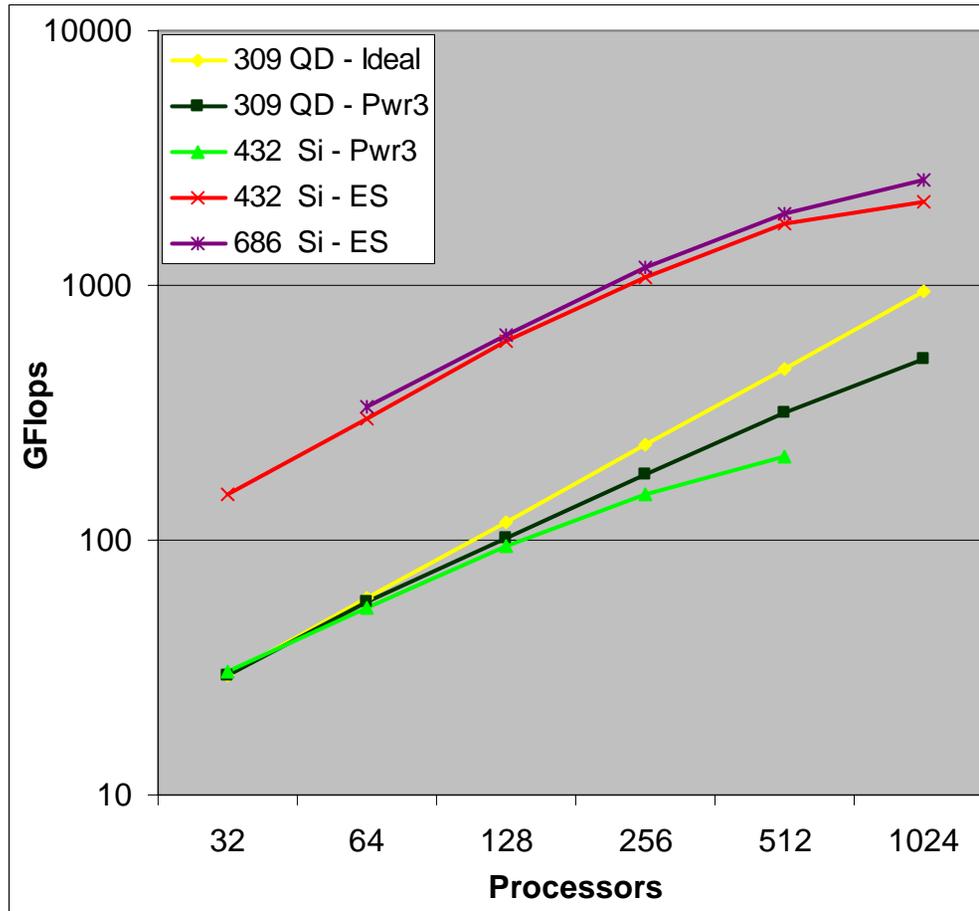
- ES can run the same system about 10 times faster than the IBM SP (on any number of processors)

- Main advantage of ES for these types of codes is the fast communication network

- Fast processors require less fine-grain parallelism in code to get same performance as RISC machines

- QD is 309 atom CdSe Quantum Dot
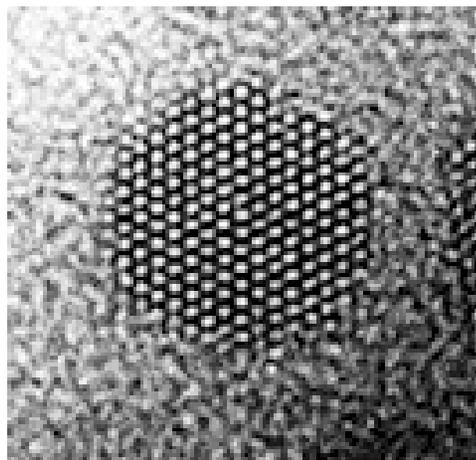
CdSe
quantum dot
TEM image
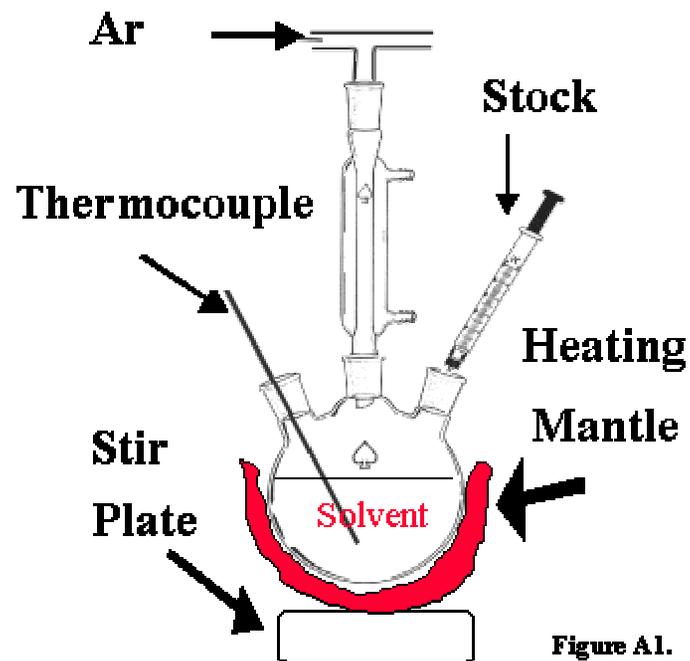
Figure A1.

- Chemically synthesised (Alivisatos, UCB, LBNL)
- Interior atoms are in bulk crystal structure
- Surface atoms are passivated
- Diameter ~ 20-100 A
- A few thousand atoms, beyond ab initio method
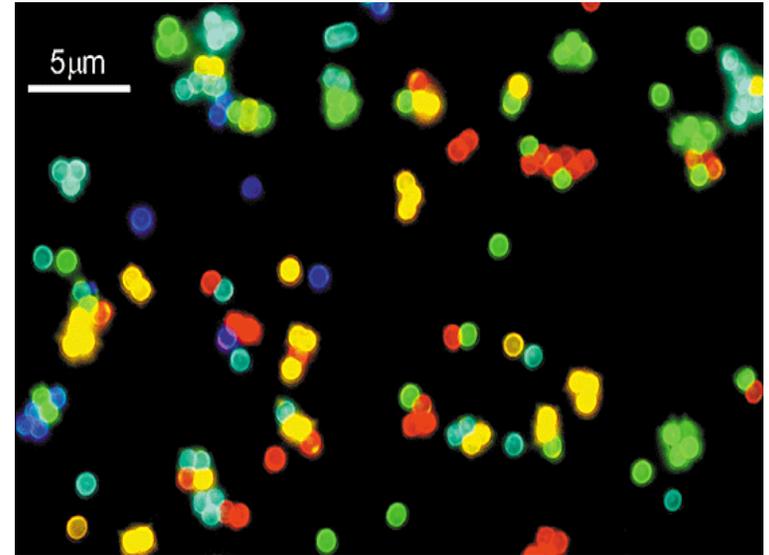
- **Optically more stable than dye molecules**
- **Can have multiple colors**

atoms molecules → **nanostructures** → **bulk**

size
- 1-100 atoms
- 1000-10^6 atoms
- Infinite (1-10 atoms in a unit cell)

method
- Ab initio Method PARATEC $O(N^3)$
- Challenge for computational nanoscience.
- • Ab initio method PARATEC
  • Effective mass method

Ab initio elements and reliability

New methodology and algorithm (ESCAN)

Even larger Supercomputer
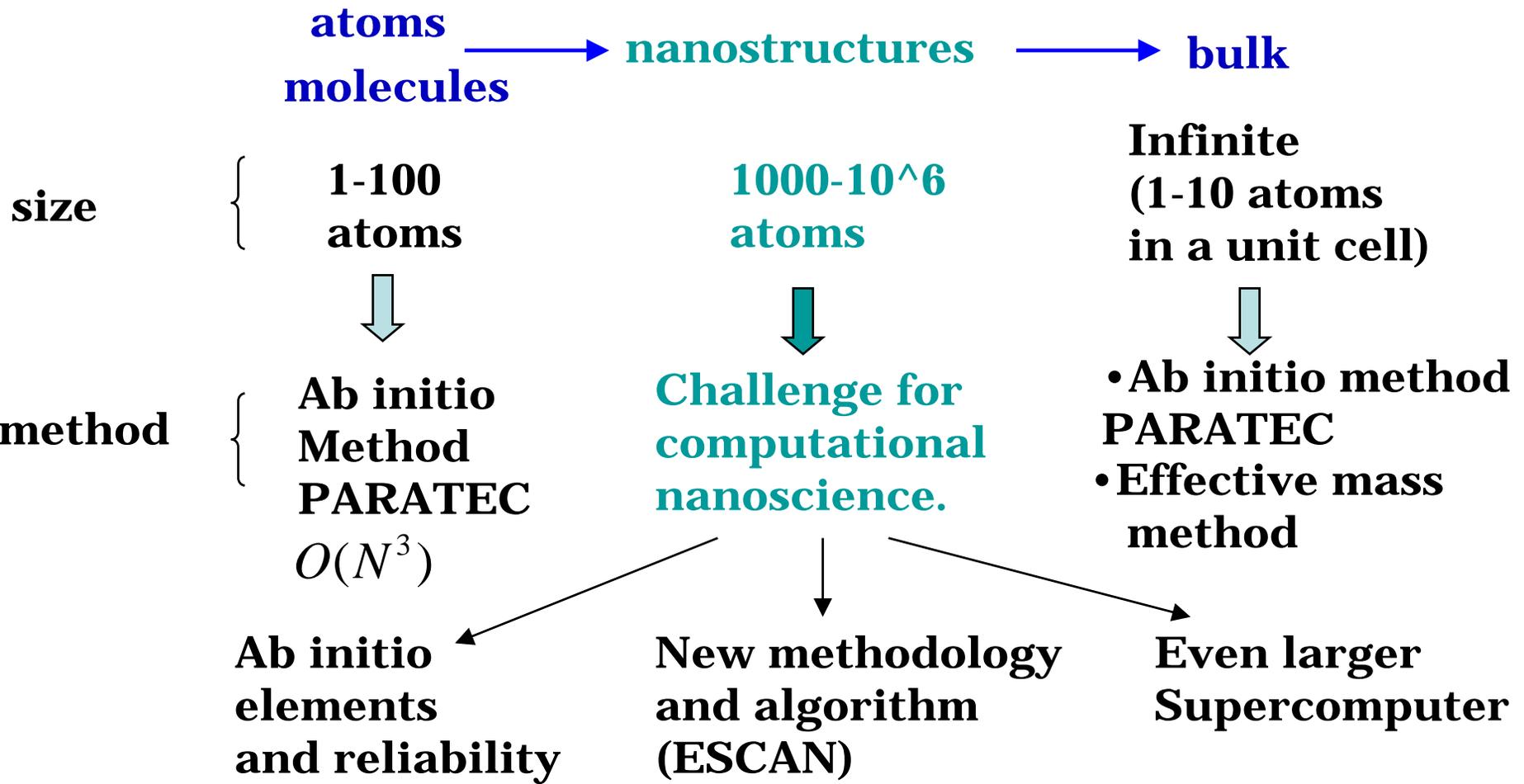
Selfconsistent LDA calculation of a single graphite sheet
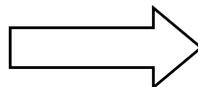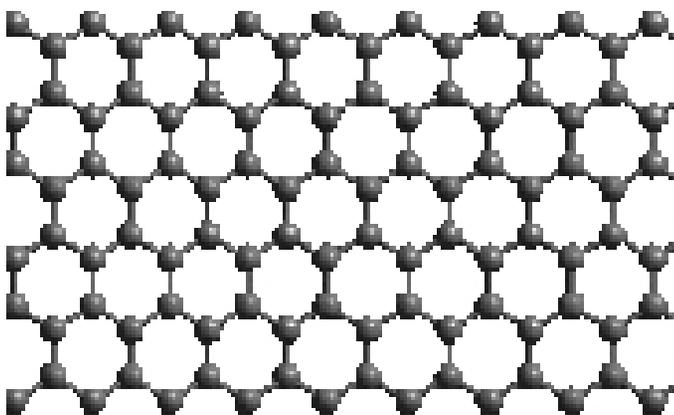
Non-selfconsistent LDA quality potential for nanotube
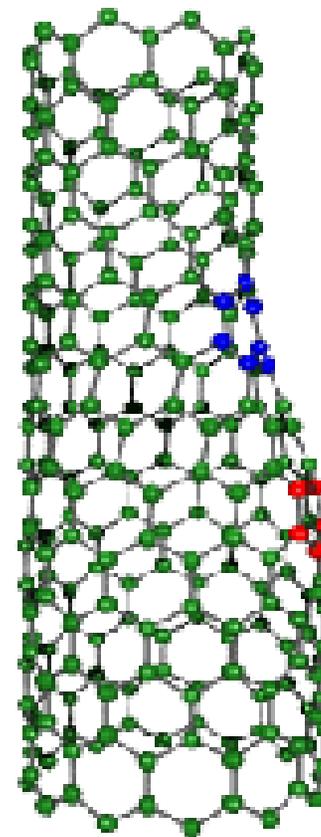
Get information from small system ab initio calc., then generate the charge densities for large systems

Office of Science
U.S. DEPARTMENT OF ENERGY

CRD

$$\{-\frac{1}{2}\nabla^2 + \int \frac{\rho(r')}{|r-r'|}dr' + \sum_I \frac{Z}{|r-R_I|} + V_{XC}(r)\}\psi_j(r) = E_j\psi_j(r)$$

Solve Kohn-Sham Equations non-selfconsistently for electron wavefunctions in desired energy range using patched charge density (can study larger nanosystems 10,000 atoms)

1. **Plane-wave expansion for** $\psi_{j,k}(r) = \sum_g C_g^j(k)e^{i(g+k).r}$

2. **Replace "frozen" core by a pseudopotential**

   Different parts of the Hamiltonian calculated in different spaces (fourier and real) 3d FFT used

Office of Science

U.S. DEPARTMENT OF ENERGY

CRD

$\rho_{graphite}$ （ LDA ） ρ_{motif}
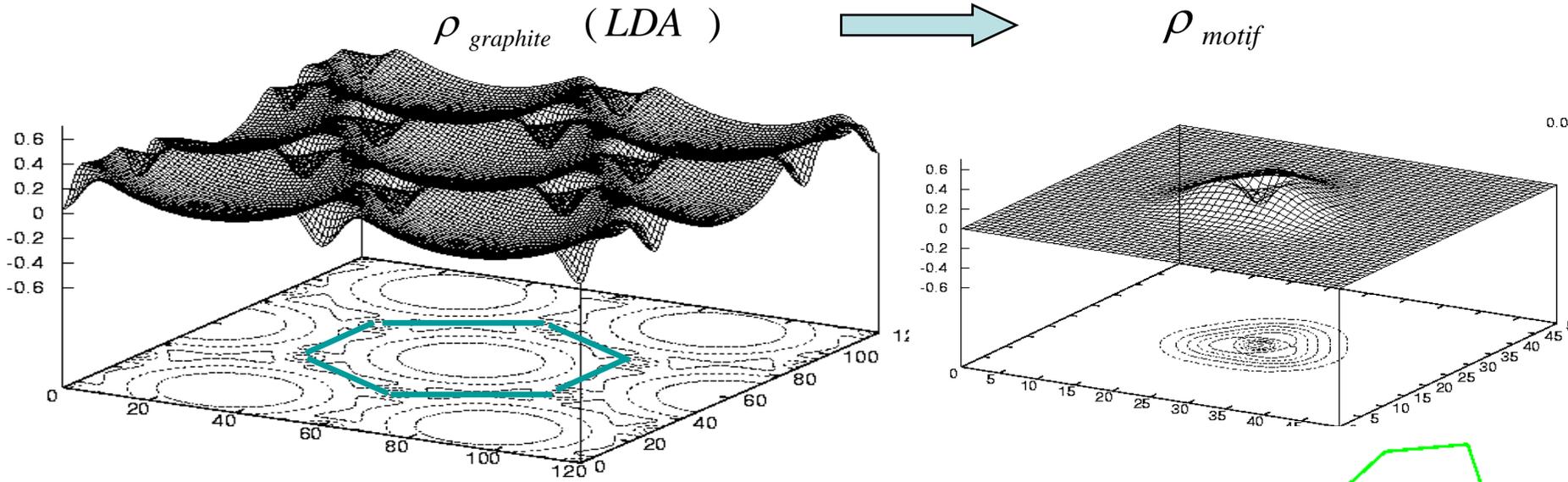
$$\rho_{nanotube}^{patch}(r) = \sum_R \rho_{motif}^{aligned}(r - R)$$

Error: 1%, ~20 meV eigen energy error.

Office of Science
U.S. DEPARTMENT OF ENERGY
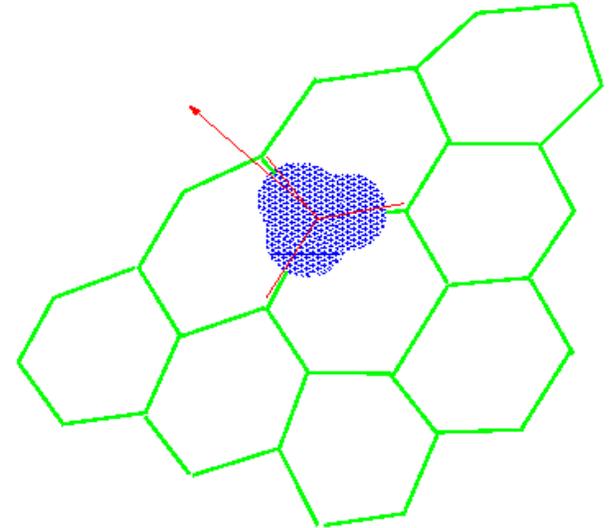
$$\{-\frac{1}{2}\nabla^2 + V(r)\}\psi_i(r) = E_i\psi_i(r)$$

$$H\psi_i = \varepsilon_i\psi_i \qquad\qquad (H - \varepsilon_{ref})^2\psi_i = (\varepsilon_i - \varepsilon_{ref})^2\psi_i$$

CBM →
ε_{ref} →
VBM →

spectrum folding

$N \{ \{\varepsilon_i\} \qquad\qquad \{ [\varepsilon_i - \varepsilon_{ref}]^2 \}$
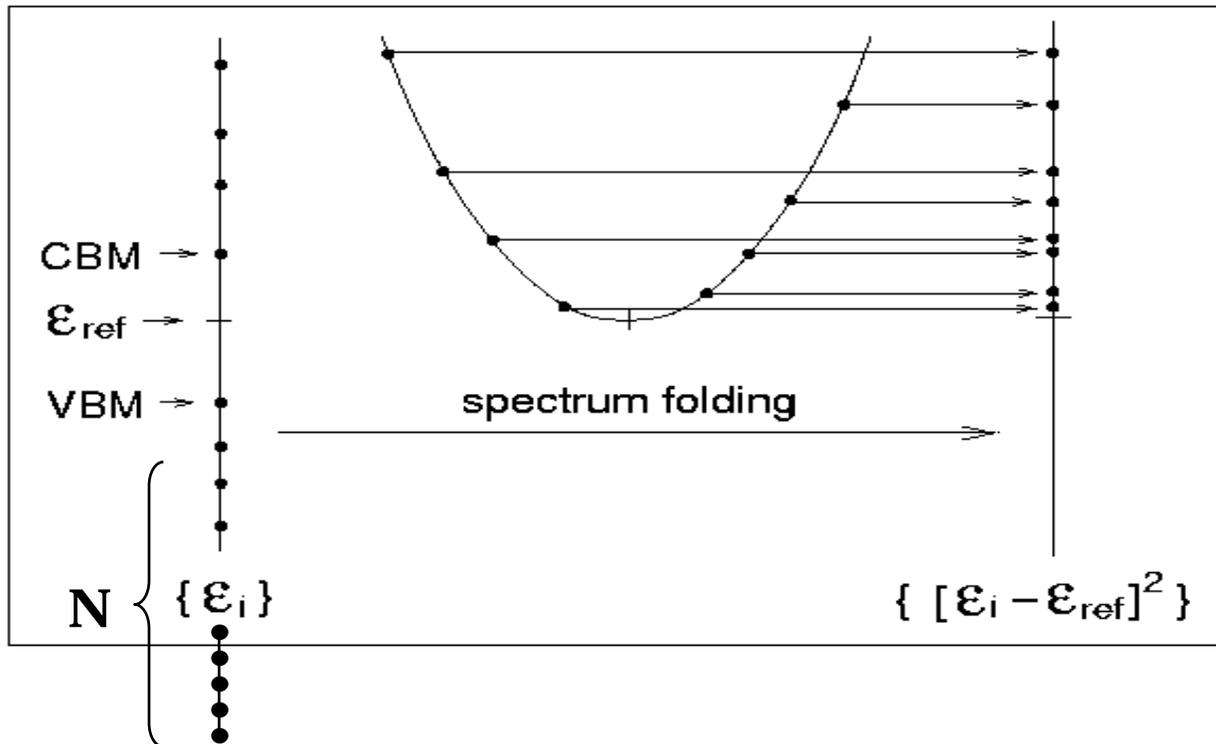
$In_{675}P_{652}$ **LDA quality calculations (eigen energy error ~ 20 meV)**

64 processors (IBM SP3) for ~ 1 hour

CBM

VBM

Total charge density motifs
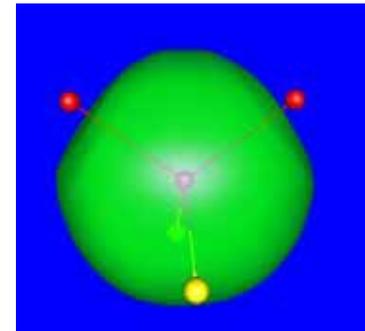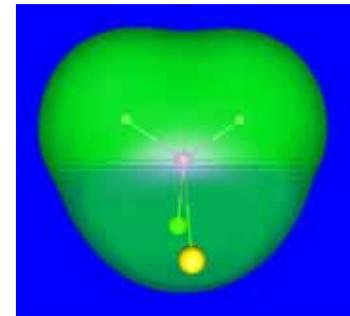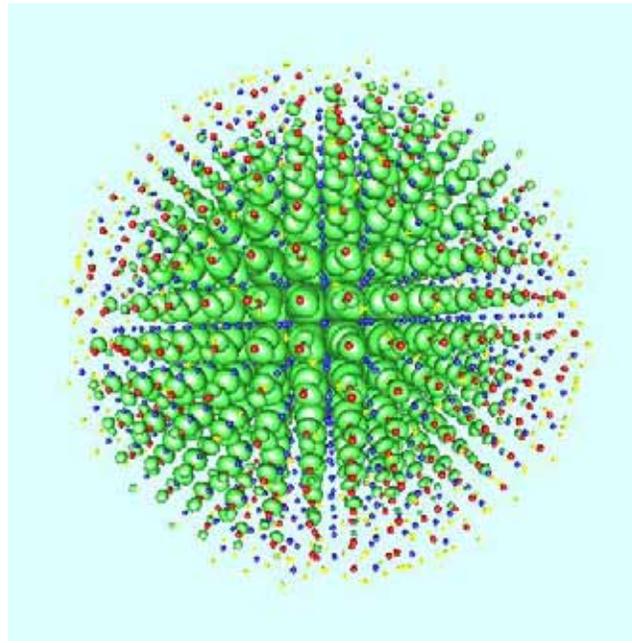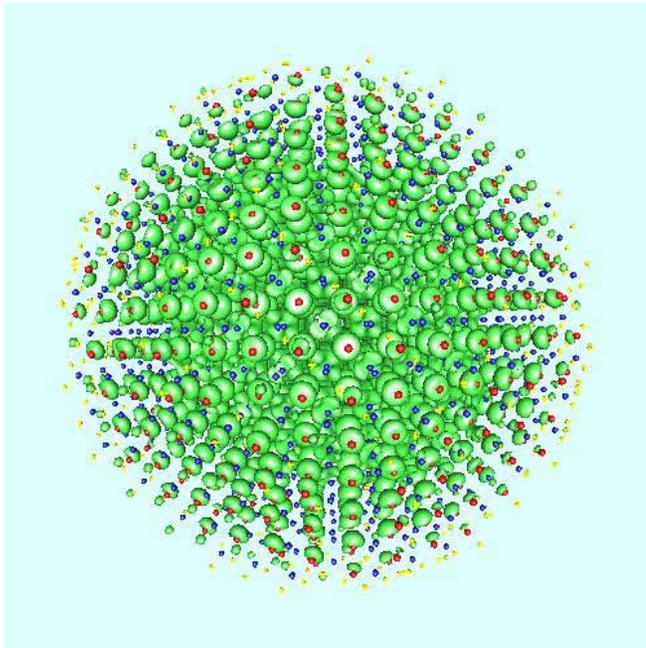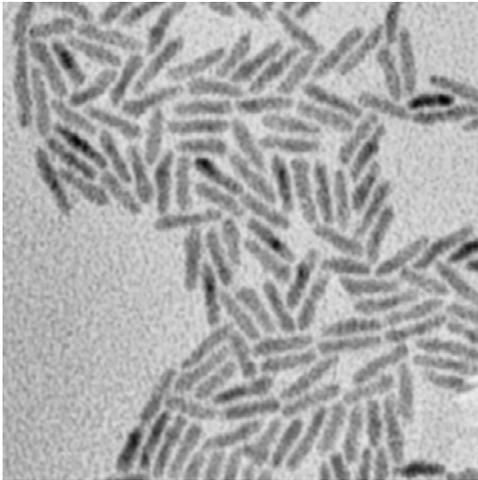
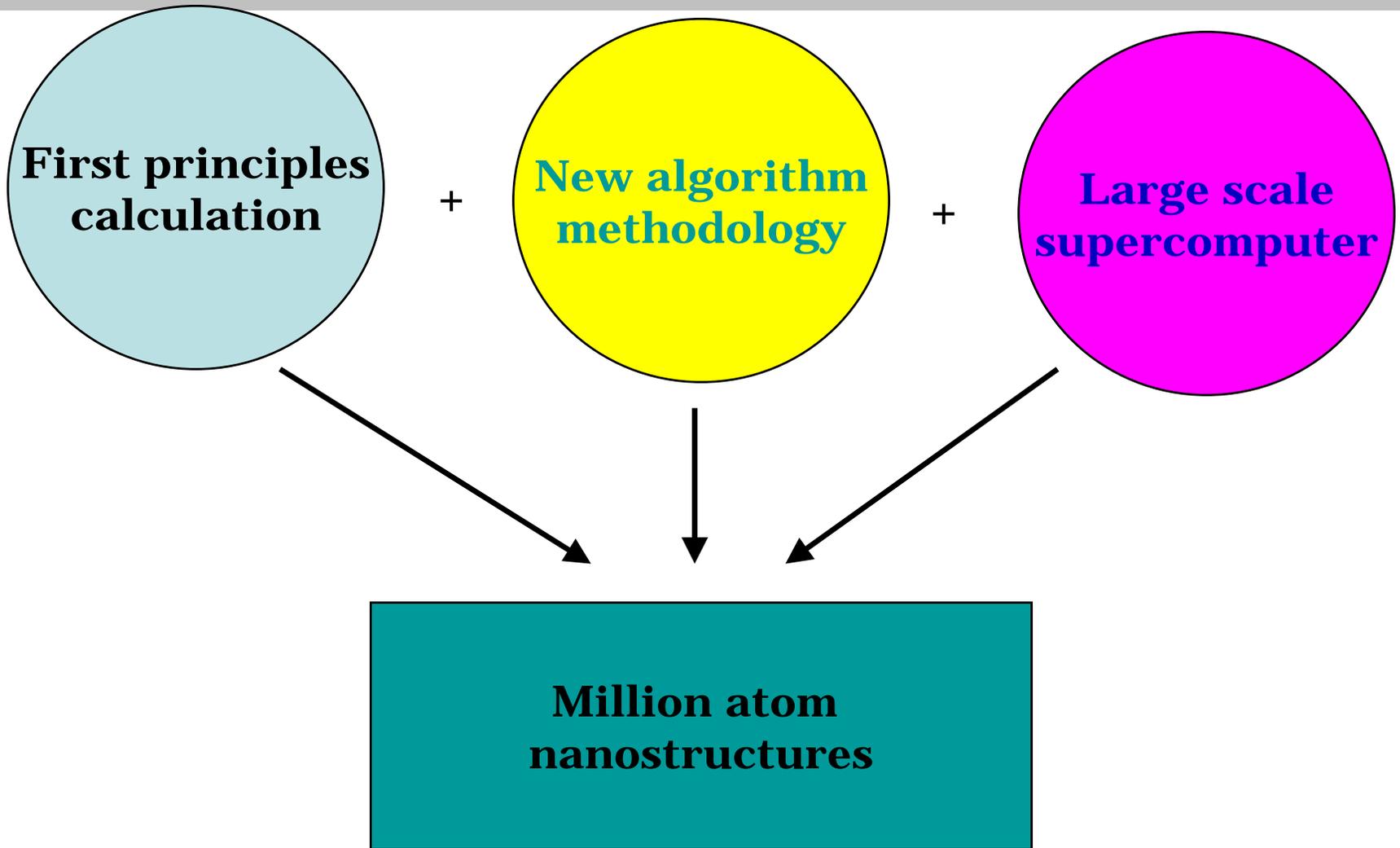**CdSe quantum rods**
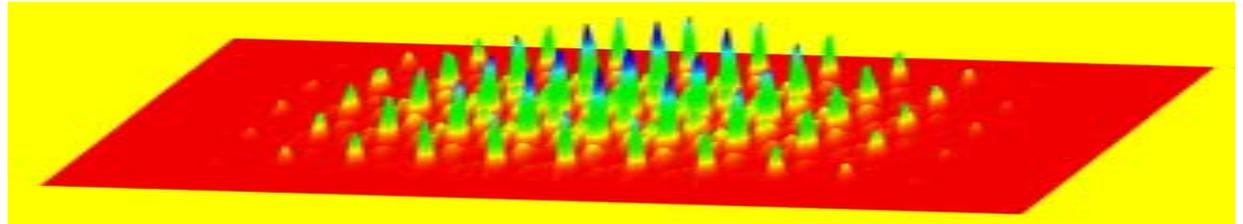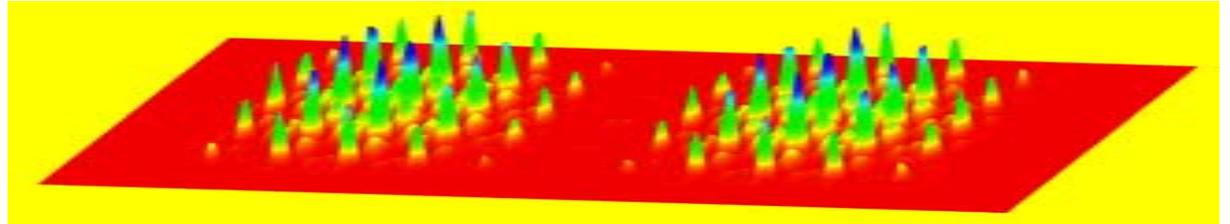
The electron wavefunctions of a quantum rods

(111) GaN wire

(112) GaN wire

CB1

CB2

Acknowledgements